Percolation in self-similar networks

M. Ángeles Serrano,[1] Dmitri Krioukov,[2] and Marián Boguñá[3]

[1]*Departament de Química Física, Universitat de Barcelona, Martí i Franquès 1, 08028, Barcelona, Spain*
[2]*Cooperative Association for Internet Data Analysis (CAIDA), University of California,*
San Diego (UCSD), 9500 Gilman Drive, La Jolla, CA 92093, USA
[3]*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*
(Dated: January 28, 2011)

We provide a simple proof that graphs in a general class of self-similar networks have zero percolation threshold. The considered self-similar networks include random scale-free graphs with given expected node degrees and zero clustering, scale-free graphs with finite clustering and metric structure, growing scale-free networks, and many real networks. The proof and the derivation of the giant component size do not require the assumption that networks are treelike. Our results rely only on the observation that self-similar networks possess a hierarchy of nested subgraphs whose average degree grows with their depth in the hierarchy. We conjecture that this property is pivotal for percolation in networks.

Percolation is a fundamental phenomenon in nature. Recent developments in percolation theory [1] open new perspectives in many areas of statistical mechanics and quantum field theory [2]. In statistical mechanics of complex networks, the percolation properties of a network determine its robustness with respect to structural damage, and dictate how emergent phenomena depend on the network structure [3]. Large clusters of connected nodes emerge above a critical value of some network parameter, e.g., the average degree; below the threshold, networks decompose into a myriad of small components. This percolation threshold can be zero, meaning that networks are always in the percolated phase. A classic example is random scale-free networks with the power-law degree distribution exponent γ lying between 2 and 3 [4, 5]. The value of the percolation threshold, the size of the giant component, and the specifics of the percolation transition strongly depend on fine details of the network topology [3]. This dependency hinders attempts to define percolation universality classes, even though some networks show some degree of percolation universality [6].

This problem is aggravated by difficulties in the analytic treatment of percolation properties for networks with strong clustering. A majority of the obtained analytic results use the generating function formalism based on the assumption that networks are locally treelike [7]. This assumption allows one to employ convenient tools from the theory of random branching processes. The assumed absence of loops implies, in particular, that clustering is zero in the thermodynamic limit. This zero-clustering approximation is valid for weakly clustered networks where triangles do not overlap, but it is invalid for networks with strong clustering and overlapping triangles observed in many real systems [8]. Noticeably, the exact results derived for some network models with clustering can be mapped to treelike zero-clustering graphs after appropriate transformations [9].

In this Letter, we provide a remarkably simple rigorous proof for the absence of a percolation threshold in a general class of self-similar networks. The proof does not rely on the treelike assumption or on generating functions. It does not depend on whether a network is weakly or strongly clustered, and it applies equally well to equilibrium or non-equilibrium networks. The proof relies only on network self-similarity, defined as statistical invariance of a hierarchy of nested subgraphs with respect to a network renormalization procedure. The percolation threshold is zero as soon as the average degree in subgraphs is a growing function of their depth in the hierarchy—a property characterizing many real networks. We also calculate analytically the size of the giant component, supporting all the results by large-scale numerical simulations.

Let $\mathcal{G}(\{\alpha\})$ be an ensemble of sparse graphs in the thermodynamic limit, where $\{\alpha\}$ is the set of model parameters. In the case of classical random graphs, for example, set $\{\alpha\}$ is just the average degree $\langle k \rangle$. Consider a transformation rule T that for each graph $G \in \mathcal{G}(\{\alpha\})$ selects one of G's subgraphs. Denote the ensemble of these subgraphs by $\mathcal{G}_T(\{\alpha\})$. The ensemble $\mathcal{G}(\{\alpha\})$ is called self-similar with respect to the transformation rule T if the transformed ensemble is the same as the original one except for some transformation of the model parameters,

$$\mathcal{G}_T(\{\alpha\}) = \mathcal{G}(\{\alpha_T\}). \qquad (1)$$

In what follows we describe three general types of graphs to which this definition applies. The first two types are equilibrium random scale-free graph ensembles belonging to a general class of network models with hidden variables [10]. The third one is a non-equilibrium ensemble of growing networks.

Type I: The graphs in this ensemble are constructed by assigning to each node a hidden variable κ drawn from the power-law probability density $\rho(\kappa) = (\gamma-1)\kappa_0^{\gamma-1}\kappa^{-\gamma}$. Without loss of generality, κ_0 can be selected such that

$\kappa \geq \kappa_0$ is the expected degree of nodes with hidden variable κ, so that the degree distribution scales as a power law with exponent γ. Each pair of nodes with expected degrees κ and κ' is then connected with probability $r(\kappa, \kappa') = f(\mu \kappa \kappa')$, where constant μ fixes the average degree $\langle k \rangle$ in the constructed graphs, and function $f(x) \leq 1$ is an arbitrary analytic function with $f(0) = 0$. This type of graphs includes as particular cases the maximally random graphs with a given expected degree sequence [11], and random graphs with arbitrary structural correlations [12]. In the former case,

$$f(x) = \frac{1}{1 + 1/x}. \tag{2}$$

Clustering vanishes in the thermodynamic limit, and therefore the treelike assumption holds.

Type II: Besides having assigned expected degrees κ, nodes in this type of graphs are also uniformly distributed in a homogeneous and isotropic D-dimensional metric space [13]. Here, we consider a circle of radius R with a constant density of nodes $\delta = N/(2\pi R)$, although all the following results can be extended to an arbitrary dimension. The connection probability between a pair of nodes with hidden variables κ and κ' separated by distance $d = (\pi - |\pi - |\theta - \theta'||)R$ on the circle (θ's are the node angular coordinates) must be of the form $r(\kappa, \theta; \kappa', \theta') = h\left(\frac{d}{\mu \kappa \kappa'}\right)$, where function h must be integrable. These graphs have the same degree distribution as the type I graphs, but clustering is finite in the thermodynamic limit, thanks to the triangle inequality in the underlying metric space [13]. Therefore the treelike assumption does not hold.

The graph sparsity in the thermodynamic limit defines constant μ in the two cases as

$$\mu_I = \frac{\langle k \rangle}{N f'(0) \kappa_0^2} \left(\frac{\gamma - 2}{\gamma - 1}\right)^2, \quad \mu_{II} = \frac{\langle k \rangle}{2\delta I \kappa_0^2} \left(\frac{\gamma - 2}{\gamma - 1}\right)^2, \tag{3}$$

where $I = \int_0^\infty h(x)dx$. Since κ_0 and δ are dumb parameters, we see that unless functions f and h contain some additional parameters, the described two graph ensembles have only two independent parameters: the power-law exponent γ and the average degree $\langle k \rangle$.

Consider now transformation rule T which simply removes all nodes with hidden variable $\kappa < \kappa_T$ from a given graph G in any of the two ensembles, where κ_T is some predefined threshold. This transformation maps the original graph G to its subgraph G_T of size $N_T = N(\kappa_0/\kappa_T)^{\gamma-1}$. The hidden variables κ of nodes remaining in G_T are distributed according to $\rho_T(\kappa) = (\gamma - 1)\kappa_T^{\gamma-1}\kappa^{-\gamma}$ with $\kappa \geq \kappa_T$. That is, the power-law exponent in G_T is the same as in G, $\gamma_T = \gamma$. The transformation does not affect the hidden variables of the nodes in subgraph G_T. Therefore the connection probability in G_T is exactly the same as in the original graph G, which

means that the ensemble of transformed graphs is identical to the ensemble of original graphs, except that the average degree has changed. Specifically, the transformation of parameters $\{\alpha\} \to \{\alpha_T\}$ in the self-similarity definition in Eq. (1) is

$$\gamma \to \gamma_T = \gamma, \quad \langle k \rangle \to \langle k \rangle_T = \langle k \rangle \left(\frac{N}{N_T}\right)^{\frac{3-\gamma}{\gamma-1}}, \tag{4}$$

which is the same for both type I and type II graphs [13]. Therefore, both ensembles belong to the same self-similarity universality class.

Type III: As opposed to the first two equilibrium ensembles, the graphs of this type are grown by adding nodes one by one. Each node i brings m_i new links, where $m_i = m_0(N/i)^\eta$ and $\eta \in [0, 1)$. Each link is then attached to a random existing node. A network is initialized with $10m_0 N^\eta$ uncounted disconnected nodes. If $\eta = 0$, the generated graphs have an exponential degree distribution. If $\eta > 0$, the degree distribution is $P(k) = (1 + m_0/\eta)^{1/\eta}/[\eta(k + m_0/\eta)^{1+1/\eta}]$, i.e., approximately a power law with exponent $\gamma = 1 + 1/\eta$. The transformation rule T simply extracts from a grown graph its subgraph composed of the first N_T nodes. With this T, the graphs of this type are also self-similar, and the parameter transformation in definition (1) is

$$\gamma \to \gamma_T = \gamma, \quad \langle k \rangle \to \langle k \rangle_T = \langle k \rangle \left(\frac{N}{N_T}\right)^{\frac{1}{\gamma-1}}. \tag{5}$$

As we show next, self-similarity of the considered ensembles (types I, II, and III), and the proportionality $N_T \sim N$ are sufficient to prove the absence of a percolation threshold for equilibrium scale-free graphs with exponent $\gamma < 3$ and for growing graphs with any γ. The key property which we will use is that the average degree of these self-similar subgraphs is a growing function of the subgraph depth in the nested subgraph hierarchy, meaning that $\langle k \rangle_T$ grows as N_T decreases in Eqs. (4,5). The same property characterizes many real networks as shown in Fig. 1 and in [13, 14].

The proof is by contradiction. As usual [4], let the average degree be the order parameter for a percolation transition. Suppose that the considered self-similar ensembles do have a non-zero percolation threshold at some critical value of the average degree $\langle k \rangle_c$. Consider a graph with the average degree below the threshold ($\langle k \rangle < \langle k \rangle_c$) which has no giant component. Since its subgraphs belong to the same ensemble, their percolation threshold is also $\langle k \rangle_c$. But since their average degree increases with their depth in the subgraph hierarchy, there exist deep enough subgraphs whose average degree is above the threshold ($\langle k \rangle_T > \langle k \rangle_c$). We thus arrive at a contradiction since a graph which does not have a giant component must contain subgraphs which do have giant components.

We next compute the size $g(p)$ of the giant component in bond percolations with bond occupation probability p,
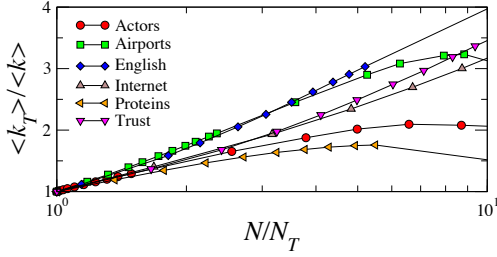
FIG. 1: Ratio of the subgraph average degree $\langle k \rangle_T$ to the average degree $\langle k \rangle$ in the whole graph as a function of the inverse relative subgraph size N/N_T for few real networks. The subgraphs are obtained by removing nodes with degrees below thresholds k_T from the original network. To insulate against finite-size effects, the data is shown only for subgraphs of size $N_T/N > 0.1$. *Actors*, actor collaborations from the Internet Movie Database; *Airports*, USA airport network; *English*, web of semantic associations between words in English; *Internet*, topology of the Internet at the Autonomous Systems level; *Proteins*, protein interaction network of *Saccharomyces cerevisiae*; and *Trust*, mutual trust relationships among individuals extracted from the Pretty Good Privacy data.

confirming the absence of the percolation threshold in the considered ensembles. We first focus on the equilibrium networks of types I and II, in which case bond percolation is equivalent to replacing the connection probability r_{ij} with pr_{ij}. Given a node i and a set of other nodes Υ, the probability that i is connected to at least one node in Υ is one minus the probability that i is not connected to any node in Υ, *i.e.*, $1 - \exp\left[\sum_{j \in \Upsilon} \ln\left(1 - pr_{ij}\right)\right]$. Node i belongs to the giant component if and only if it is connected to the giant component of the graph without i. If $\widetilde{g}_j(p)$ denotes the probability that this i-deprived component contains some other node j, then

$$g_i(p) = 1 - \exp\left[\sum_{j \neq i} \widetilde{g}_j(p) \ln\left(1 - pr_{ij}\right)\right]. \quad (6)$$

Since in small-world networks a single node cannot significantly affect the percolation properties of the rest of the graph, we identify $\widetilde{g}_j(p) = g_j(p)$, transforming Eq. (6) into a self-consistent equation for $g_i(p)$. We note that Eq. (6) does not use the treelike assumption. This equation is thus valid for the type II graphs with strong clustering as well as for zero-clustering type I graphs. It leads in the thermodynamic limit to the following expression for the probability $g(\kappa; p)$ that a node with expected degree κ belongs to the giant component:

$$g(\kappa; p) = 1 - e^{-\kappa\psi(p)}, \quad \text{where } \psi(p) \text{ satisfies} \quad (7)$$

$$\frac{[\psi(p)]^{3-\gamma}}{a(p)} = \frac{[\psi(p)]^{2-\gamma}}{\gamma - 2} - \Gamma[2 - \gamma, \psi(p)], \quad \text{with} \quad (8)$$

$$a_I(p) = \frac{(\gamma - 2)^2}{\gamma - 1}\langle k \rangle p, \quad (9)$$

$$a_{II}(p) = -\frac{(\gamma - 2)^2}{\gamma - 1}\frac{\langle k \rangle}{I}\int_0^\infty \ln\left(1 - ph(x)\right)dx \quad (10)$$
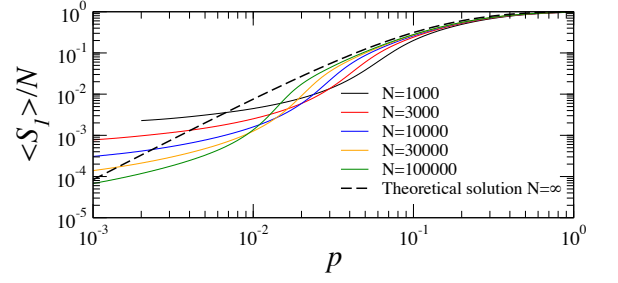

FIG. 2: Relative size of the largest cluster $\langle S_1 \rangle/N$ vs. the analytical solution for type II networks with $\gamma = 2.5$, $\langle k \rangle = 3(\gamma-1)/(\gamma-2)$, $h(x) = e^{-x}$, and average clustering coefficient (measured over degrees larger than $k = 1$) $\bar{c} = 0.5$.

for types I and II, respectively. The size of the giant component is then $g(p) = \int \rho(\kappa)g(\kappa; p)d\kappa = 1 - (\gamma - 1)E_\gamma[\psi(p)]$, where E_γ is the extended exponential integral. In diluted networks with $p \ll 1$, $a_I(p) \approx a_{II}(p)$, and the giant component size for both classes becomes

$$g(p) \sim \left[-\frac{(\gamma - 2)^{\gamma-1}}{(\gamma - 1)^{\gamma-2}\Gamma(2 - \gamma)}\langle k \rangle p\right]^{\frac{1}{3-\gamma}}. \quad (11)$$

The value of the critical exponent β in $g(p) \sim p^\beta$ is thus $\beta = 1/(3 - \gamma)$, agreeing with [15]. We emphasize that in our case, this result is obtained without using the treelike assumption. Therefore, quite surprisingly, this exponent characterizes equilibrium scale-free networks with arbitrary clustering and degree correlations.

In non-equilibrium networks of type III, we can compute an upper bound for β. Self-similarity of these networks, coupled with the observation that any node belonging to the giant component of a self-similar subgraph of a type III graph belongs also to the giant component of the graph itself, leads to inequality

$$g(p) \geq \frac{N_T}{N}g\left(\left[\frac{N}{N_T}\right]^\eta p\right). \quad (12)$$

By choosing $N_T/N = p^{1/\eta}$, we obtain $g(p) \geq p^{1/\eta}g(1)$. Therefore the exponent β satisfies $\beta \leq 1/\eta = \gamma - 1$. We see that growth reduces significantly this exponent, compared to the equilibrium case with the same γ.

We next check our analytic results against large-scale simulations. We generate type I and II networks using the connection probabilities in Eq. (2) and $h(x) = e^{-x}$, respectively. We do not allow κ's above the natural cutoff $\kappa_c = N^{1/(\gamma-1)}$. For all the three graph types, for each graph size N ranging from 10^3 to 10^5, and for each value of the bond occupation probability p, we generate 10^3 graphs, and for each graph we perform bond percolation 10^4 times. For each percolation we measure the size S_1 of the largest connected component in the graph using the fast algorithm of Newman and Ziff [16], and calculate the average $\langle S_1 \rangle$ of the largest component size and its fluctuations, i.e., susceptibility $\chi = \sqrt{\langle (S_1 - \langle S_1 \rangle)^2 \rangle}$, for each combination of p, N, and graph type.
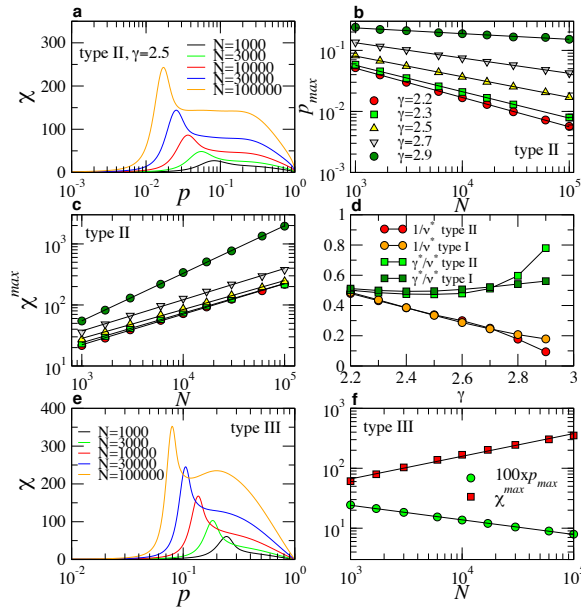
FIG. 3: Bond percolation simulations for equilibrium (types I and II) and growing (type III) networks . **a:** susceptibility χ as a function of bond occupation probability p and graph size N for the same network as in Fig. 2. **b** and **c:** position p_{max} and height χ^{max} of the peak of χ as functions of network size N. The straight lines are power law fits. **d:** exponents $1/\nu^*$ and γ^*/ν^* in $p_{max}(N) \sim N^{-1/\nu^*}$ and $\chi^{max}(N) \sim N^{\gamma^*/\nu^*}$ for the type I and II graphs. **e** and **f:** Bond percolation simulations for non-equilibrium networks (type III) with $\eta = 1/4$ ($\gamma = 5$) and $m_0 = 2$. The measured values of the scaling exponents are $1/\nu^* = 0.24(3)$ and $\gamma^*/\nu^* = 0.3(8)$.

Since the convergence to the thermodynamic limit in scale-free networks is slow [17], it is difficult to accurately measure exponent β in simulations. Nevertheless we observe an agreement, albeit slowly converging, between the analytical solution for $g(p)$ and simulations in Fig. 2. In Fig. 3 we also show susceptibility $\chi(p, N)$ for equilibrium and growing networks. Susceptibility displays peaks whose positions $p_{max}(N)$ and heights $\chi^{max}(N)$ depend as power laws on the system size, $p_{max}(N) \sim N^{-1/\nu^*}$ and $\chi^{max}(N) \sim N^{\gamma^*/\nu^*}$. Taken together, these two results confirm that the giant component emerges at $p = p_{max}$, and that the percolation threshold vanishes in the thermodynamic limit $N \to \infty$ where $p_{max} \to 0$ and $\chi^{max} \to \infty$.

In short, self-similar networks with subgraphs of growing average degree have no percolation threshold. The proof can be generalized to any processes with phase transitions whose critical points depend monotonously on the average degree. Examples include, among others, the absence of an epidemic threshold in epidemic spreading processes, or the absence of a paramagnetic phase in the Ising model on scale-free networks [3].

The identification of percolation universality classes for general random networks is a notoriously difficult problem—details tend to prevail. Nevertheless, the re-

sults presented here lead us to conjecture that *self-similar* networks can be split into three general percolation universality classes, depending only on whether the average degree in the nested subgraph hierarchy increases, remains constant, or decreases with the subgraph depth, and independent of any other network properties, such as clustering, correlations, equilibrium vs. non-equilibrium classification, etc.

[1] S. Smirnov, C. R. Acad. Sci. Paris Sr. I Math **333**, 239 (2001).
[2] J. Cardy, Ann. Henri Poincaré **4, Suppl. 1**, S371 (2003).
[3] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[4] M. Molloy and B. Reed, Random Structures and Algorithms **6**, 161 (1995).
[5] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000).
[6] P. L. Krapivsky and B. Derrida, Physica A **340**, 714 (2004).
[7] A. Vázquez and Y. Moreno, Physical Review E **67**, 015101(R) (2003); A. V. Goltsev, S. N. Dorogovtsev, and J. F. F. Mendes, Phys. Rev. E **78**, 051105 (2008).
[8] M. A. Serrano and M. Boguñá, Phys. Rev. E **74**, 056114 (2006).
[9] M. E. J. Newman, Phys Rev E **68**, 026121 (2003); M. A. Serrano and M. Boguñá, Phys Rev E **74**, 056115 (2006); M. A. Serrano and M. Boguñá, Phys. Rev. Lett. **97**, 088701 (2006); M. E. J. Newman, Phys. Rev. Lett. **103**, 058701 (2009); J. P. Gleeson, Phys. Rev. E **80**, 036107 (2009).
[10] M. Boguñá and R. Pastor-Satorras, Phys Rev E **68**, 036112 (2003).
[11] J. Park and M. E. J. Newman, Phys Rev E **68**, 026112 (2003); D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **78**, 015101 (2008).
[12] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).
[13] M. Á. Serrano, D. Krioukov, and M. Boguñá, Phys Rev Lett **100**, 078701 (2008).
[14] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, Networks and Heterogeneous Media **3**, 371 (2008).
[15] R. Cohen, D. ben Avraham, and S. Havlin, Phys. Rev. E **66**, 036113 (2002); M. Ostilli and J. F. F. Mendes, Eur. Phys. Lett. **92**, 40013 (2010).
[16] M. E. J. Newman and R. M. Ziff, Phys. Rev. Lett. **85**, 4104 (2000).
[17] M. Boguñá, C. Castellano, and R. Pastor-Satorras, Phys. Rev. E **79**, 036110 (2009).